



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 21/04, C07K 14/705, C12N 15/09, 15/63, C12Q 1/68		A1	(11) International Publication Number: WO 99/64436 (43) International Publication Date: 16 December 1999 (16.12.99)
(21) International Application Number: PCT/US99/12773 (22) International Filing Date: 8 June 1999 (08.06.99) (30) Priority Data: 60/089,098 12 June 1998 (12.06.98) US (71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): FEIGHNER, Scott, D. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). PATCHETT, Arthur, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). TAN, Carina [MY/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). MC-KEE, Karen [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). MACNEIL, Douglas [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). HOWARD, Andrew, D. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). PONG, Sheng-Shung [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). SMITH, Roy, G. [GB/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: CLONING AND IDENTIFICATION OF THE MOTILIN RECEPTOR			
(57) Abstract The motilin receptor has been isolated and cloned, and nucleic acid sequences are given. Two splice variants have been identified. Also, assays for motilin receptor ligands are given. The identification of the cloned motilin receptor may be used to screen and identify compounds which bind to the receptor for use in a variety of gastric conditions, including gastric motility disorders.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE OF THE INVENTION
CLONING AND IDENTIFICATION OF THE MOTILIN RECEPTOR

CROSS-REFERENCE TO RELATED APPLICATIONS

5

XXXXX

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

XXXXX

10 REFERENCE TO MICROFICHE APPENDIX

XXXXX

FIELD OF THE INVENTION

15 The present invention is directed to a novel human DNA sequence encoding a motilin receptor, the receptor encoded by the DNA, and the uses thereof.

BACKGROUND OF THE INVENTION

20 Gastrointestinal (GI) motility is a coordinated neuromuscular process which transports nutrients through the digestive system. Impaired GI motility, can lead to irritable bowel syndrome, constipation and diabetic and post-surgical gastroparesis and is one of the largest health care burdens of industrialized nations. Motilin, a 22 amino acid prokinetic peptide is expressed throughout the gastrointestinal tract in a
25 number of species including humans. Released from endochromaffin cells of the small intestine, motilin exerts a profound effect on gastric motility with the induction of interdigestive (phase III) antrum and duodenal contractions. The unrelated macrolide antibiotic erythromycin also possesses prokinetic properties mediated by its
30 interaction with motilin receptors. These account for erythromycin's GI side-effects, including vomiting, nausea, diarrhea and abdominal muscular discomfort.

35 Motilin receptors have been detected in the GI tract and recently in the central nervous system, but their molecular structure has not been reported. Although motilin receptor characterization has been actively pursued in humans and other species since the isolation of motilin from

porcine intestine in 1972, the receptor itself has not been isolated nor cloned.

5 Motilin is highly conserved across species and is synthesized as part of larger pre-prohormone. Mature 22 amino acid motilin is generated by removal of its secretory signal peptide and cleavage at the first C-terminally located dibasic prohormone convertase recognition site. Using radioligand binding, autoradiography and *in vitro* bioassays, high affinity and low density, motilin receptors were detected in smooth muscle cells of the gastrointestinal tract of humans, cats and rabbits.

10 Cerebellar brain receptors for motilin were also described supporting the notion that motilin may act in the central nervous system. Native motilin receptors appear to be coupled to G proteins and activate the phospholipase C signal transduction pathway resulting in Ca^{2+} influx through L-type channels.

15 The development of safe and selective motilin receptor agonists is likely to aid the treatment of disorders resulting from impaired GI motility. Thus, it would be desirable to be able to isolate, and clone the motilin receptor, and to use this in assays for agonists and antagonists.

20 SUMMARY OF THE INVENTION

The present invention is directed to a novel G-protein coupled receptor (GPCR), designated as motilin receptor. Two spliced forms of the motilin receptor were identified: MTL-R1A, which encodes a functional seven-transmembrane domain form, and MTL-R1B, which encodes a truncated five-transmembrane domain form.

25 Both forms make up embodiments of this invention.

Another aspect of this invention are nucleic acids which encode the motilin receptor, which are isolated, or free from associated nucleic acids.

30 Other aspects of this invention include assays for identifying motilin ligands which are agonists and antagonists of a motilin receptor comprising contacting a candidate ligand with a motilin receptor and determining if binding occurred.

Another aspect of this invention is a method for

35 determining whether a ligand is capable of binding to a motilin receptor comprising:

- (a) transfecting test cells with an expression vector encoding motilin receptor;
- (b) exposing the test cells to the ligand;
- (c) measuring the amount of binding of the ligand to the motilin receptor;
- (d) comparing the amount of binding of the ligand to the motilin receptor in the test cells with the amount of binding of the ligand to control cells that have not been transfected with the motilin receptor
- where if the amount of binding of the ligand to the test cells is greater than the amount of binding of the ligand to the control cells, then the substance is capable of binding to motilin receptor.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the DNA sequence of motilin receptor gene including 5' untranslated region (SEQ.ID.NO.:1). Intronic sequences are shown in lower case type.

Figure 2 shows the DNA sequence of motilin receptor spliced form A (MTL-R1A) (SEQ.ID.NO.:2).

Figure 3 shows deduced amino acid sequence of MTL-R1A (SEQ.ID.NO.:3).

Figure 4 shows the DNA sequence of motilin receptor spliced form B (MTL-R1B) (SEQ.ID.NO.:4).

Figure 5 shows the deduced amino acid sequence of MTL-R1B (SEQ.ID.NO.:5).

Figures 6 A-C compare DNA and protein sequence for MTL-R1A and MTL-R1B.

Figure 7 shows the DNA sequence of puffer fish clone 75E7 (SEQ.ID.NO.:6).

Figure 8 shows the deduced amino acid sequence of puffer fish clone 75E7 protein sequences (SEQ.ID.NO.:7).

Figure 9 shows the comparison of human MTL-R1A and puffer fish clone 75E7 protein sequences.

Figure 10 is a graph illustrating the pharmacological characterization of the cloned MTL-R1A in the aequorin bioluminescence assay in HEK-293 cells.

Figure 11 is a graph illustrating the pharmacological characterization of the cloned MTL-R1A in the [¹²⁵I]-Tyr⁷-human motilin binding assay.

5 As used throughout the specification and claims, the following definitions apply:

 “Substantially free from other proteins” means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other proteins. Thus, for example, a MTL-R1 protein
10 preparation that is substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non- MTL-R1 proteins. Whether a given MTL-R1 protein preparation is substantially free from other proteins can be
15 determined by such conventional techniques of assessing protein purity as, *e.g.*, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) combined with appropriate detection methods, *e.g.*, silver staining or immunoblotting.

 “Substantially free from other nucleic acids” means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other nucleic acids. Thus, for example, a MTL-R1 DNA
20 preparation that is substantially free from other nucleic acids will contain, as a percent of its total nucleic acid, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and
25 even more preferably no more than 0.1%, of non- MTL-R1 nucleic acids. Whether a given MTL-R1 DNA preparation is substantially free from other nucleic acids can be determined by such conventional techniques of assessing nucleic acid purity as, *e.g.*, agarose gel electrophoresis combined with appropriate staining methods, *e.g.*,
30 ethidium bromide staining, or by sequencing.

 “Functional equivalent” means a receptor which does not have the exact same amino acid sequence of a naturally occurring motilin receptor, due to alternative splicing, deletions, mutations, or additions, but retains at least 1%, preferably 10%, and more preferably
35 25% of the biological activity of the naturally occurring receptor. Such derivatives will have a significant homology with a motilin receptor and

can be detected by reduced stringency hybridization with a DNA sequence obtained from a motilin receptor. The nucleic acid encoding a functional equivalent has at least about 50% homology at the nucleotide level to a naturally occurring receptor nucleic acid.

5 "Ligand" means any molecule which binds to a motilin receptor of this invention. These ligands can have either agonist, partial agonist, partial antagonist or antagonist activity.

DETAILED DESCRIPTION OF THE INVENTION

10 The cloning of GPCR's related to the hypothalamic and pituitary receptor for the growth hormone (GH) secretagogues (GHSs) which mediate sustained pulsatile GH release has been recently described. (McKee *et. al.*, 1997 *Genomics* 46:426-434, which is hereby incorporated by reference). One of these clones, GPR38, possessed the
15 most significant amino acid sequence identity to the human GHSR (52%) (rising to as high as 86% in transmembrane domains (TM). GPR38 was classified as an orphan GPCR (GPCRs for which a natural ligand has not been identified).

GPR38 was isolated from a human genomic DNA library
20 and contained a single intron of approximately 1 kb, as shown in FIGURE 1. cDNA clones were isolated to obtain the nucleotide sequence of correctly spliced GPR38 mRNA. Efforts to isolate cDNA clones by standard library screening proved unsuccessful.

A combination of RACE and RT-PCR techniques resulted
25 in the identification of two spliced forms for GPR38. These two GPR38 cDNAs use distinct splice donor sites and a common acceptor site (perfect match to consensus exon-intron splice acceptor junction sequence [pyrimidine-rich stretch ag/TG]). GPR38-A mRNA (imperfect match to consensus donor sequence [TGC/gt]) encodes a polypeptide of
30 412 amino acids with seven alpha-helical TM domains, the hallmark feature of GPCR-Rs, whereas GPR38-B encodes a 363 amino acid polypeptide with five TM domains (perfect donor sequence [CCG/gt]). Northern blot analysis failed to reveal an expression profile for GPR38. However, when RNase protection was employed expression was
35 demonstrated in stomach, thyroid and bone marrow.

It accordance with this invention, it has been found that GPR38 is the motilin receptor. Thus, this invention is directed to the human motilin receptor, its functional equivalents, motilin receptors from other species which can be isolated using fragments of the human motilin DNA as probes, and to splice variants of the motilin receptor.

The intact motilin receptor of this invention was found to have structural features which are typical of G-protein linked receptors, including seven transmembrane (TM) domains, three intra- and extracellular loops, and the GPCR protein signature sequence. The TM domains and GPCR protein signature sequence are noted in the protein sequences of the GPCR in Figures 6A-C.

A high-throughput assay was developed which measures Ca^{2+} release with the bioluminescent Ca^{2+} sensitive-aequorin reporter protein (capable of measuring ligand-induced IP_3 -coupled mobilization of intracellular calcium and concomitant calcium-induced aequorin bioluminescence). Expression of cloned GPR38-A in cell membranes was confirmed using epitope-tagged protein which revealed a single protein species with a molecular weight of approximately 45,000 daltons containing an open reading frame encoding 412 amino acids (SEQ. ID.NO.:3). The DNA and deduced amino acid sequence are given in SEQ.ID. NO.:2 and SEQ.ID. NO.:3, respectively.

A broad set of peptide and non-peptide molecules were tested at a single concentration in transiently transfected HEK-293/aeq17 cells (100 nM peptide, 10 μM non-peptide). Significant bioluminescent responses were recorded for the peptide motilin and the non-peptide macrolide erythromycin, which was reported to be a competitive agonist at motilin receptors. Full dose-response curves confirmed this observation.

Nucleotide sequence analysis revealed two splice forms of human motilin receptor both of which make up further aspects of this invention. The first (MTL-R1A) encodes a seven transmembrane domain receptor. The full length open reading frame appears to contain 412 amino acids. The second splice form (MTL-R1B) diverges in its nucleotide sequence from MTL-R1A just before the predicted amino acid of the sixth transmembrane domain (TM6).

In the MTL-R1B, TM5 is truncated and fused to a contiguous reading frame of about 86 amino acids, followed by a translation stop codon. The DNA and amino acids sequences encoding MTL-R1A and MTL-R1B are given in FIGURES 2-5.

5 A further aspect of this invention is a related motilin receptor gene, evident in the teleost puffer fish *Spheroides nephelus*. Screening of a puffer fish genomic library identified a single clone (75E7) containing an open reading frame of 363 amino acids (approximately 54% identical at the protein level) which contains a
10 similar exon-intron structure to GPR38. Analysis of clone 75E7 shows an amino acid sequence to contain 363 amino acids with a molecular weight of approximately 41,323 daltons. (FIGURE 8). DNA sequence of puffer fish clone 75E7 is given in SEQ.ID.NO.:6, and a comparison of human MTL-R1A and puffer fish clone 75E7 protein sequences is
15 given in FIGURE 9.

Another aspect of this invention relates to vectors which comprise nucleic acids encoding a motilin receptor or a functional equivalent. These vectors may be comprised of DNA or RNA; for most cloning purposes DNA vectors are preferred. Typical vectors include
20 plasmids, modified viruses, bacteriophage and cosmids, yeast artificial chromosomes and other forms of episomal or integrated DNA that encode a motilin receptor. It is well within the skill of the ordinary artisan to determine an appropriate vector for a particular gene transfer or other use.

25 A further aspect of this invention are host cells which are transformed with a gene which encodes a motilin receptor or a functional equivalent. The host cell may or may not naturally express a motilin receptor on the cell membrane. Preferably, once transformed, the host cells are able to express the motilin receptor or a functional
30 equivalent on the cell membrane. Depending on the host cell, it may be desirable to adapt the DNA so that particular codons are used in order to optimize expression. Such adaptations are known in the art, and these nucleic acids are also included within the scope of this invention. Generally mammalian cell lines, such as HEK-293, COS-, CHO, HeLa,
35 NS/), CV-1, GC, GH3 or VERO cells are preferred host cells, but other

cells and cell lines such as *Xenopus oocytes* or insect cells, may also be used.

Human embryonic kidney (HEK 293) cells and Chinese hamster ovary (CHO) cells are particularly suitable for expression of motilin receptor proteins because these cells express a large number of G-proteins. Thus, it is likely that at least one of these G-proteins will be able to functionally couple the signal generated by interaction of motilin receptors and their ligands, thus transmitting this signal to downstream effectors, eventually resulting in a measurable change in some assayable component, *e.g.*, cAMP level, expression of a reporter gene, hydrolysis of inositol lipids, or intracellular Ca²⁺ levels.

A variety of mammalian expression vectors can be used to express recombinant motilin in mammalian cells. Commercially available mammalian expression vectors which are suitable include, but are not limited to, pCR2.2 (Invitrogen), pMC1neo (Stratagene), pSG5 (Stratagene), pcDNA1 and pcDNA1amp, pcDNA3, pcDNA3.1, pCR3.1 (Invitrogen), EBO-pSV2-neo (ATCC 37593), pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), and pSV2-dhfr (ATCC 37146). Following expression in recombinant cells, motilin receptors can be purified by conventional techniques to a level that is substantially free from other proteins.

The specificity of binding of compounds showing affinity for motilin receptors is shown by measuring the affinity of the compounds for recombinant cells expressing the cloned receptor or for membranes from these cells. Expression of the cloned receptor and screening for compounds that bind to motilin receptors or that inhibit the binding of a known, radiolabeled ligand of motilin receptors to these cells, or membranes prepared from these cells, provides an effective method for the rapid selection of compounds with high affinity for a motilin receptor. Such ligands need not necessarily be radiolabeled but can also be nonisotopic compounds that can be used to displace bound radiolabeled compounds or that can be used as activators in functional assays. Compounds identified by the above method are likely to be agonists or antagonists of

motilin receptors and may be peptides, proteins, or non-proteinaceous organic molecules.

Such molecules are useful in treating a variety of gastric conditions, including gastric motility disorders (intrinsic myopathies or neuropathy), functional defects, disorders which are secondary to
5 neurologic disorders including spinal cord transections, amyloidosis, collagen vascular disease (e.g. scleroderma), paraneoplastic syndromes, radiation-induced dysmotility, diabetes, infections, stress-related motility disorders, psychogenic/functional disorders,
10 other drugs which affect motility (e.g. beta andadrenergic drugs which may delay gastric emptying, cholinergic agents or opiates, or serotonin receptor antagonists), gastroparesis (diabetic or post-surgical), gastro-esophageal reflux disease, constipation, chronic idiopathis pseudo-obstruction and acute fecal impaction,
15 postoperative ileus, gallstones, infantile collic, preparation for colonoscopy and endoscopy, duodenal intubation , irritable bowel syndrome, non-ulcer dyspepsion, non-cardiac chest pain and diarrhea.

The pharmacological characterization of the cloned MTL-
20 R1A in the aequorin bioluminescence assay in HEK-293 cells is shown in Figure 10 and in the [¹²⁵I]-Tyr⁷-human motilin binding assay (Figure 11). Motilin at concentrations as high as 10 μ M gave no bioluminescent response above background levels in cells that were not transfected with the MTL-R1A cDNA expression vector. Similarly,
25 non-transfected cells did not show appreciable binding of [¹²⁵I]-Tyr⁷-human motilin.

The rank order of potency for motilin, motilin peptide fragments and non-peptide molecules is consistent with experiments performed on native motilin receptors, from stomach or intestinal
30 tissues.

Due to the high degree of homology to GPCRs, the motilin receptor of this invention is believed to function similarly to GPCRs and have similar biological activity. They are useful in understanding the biological and physiological pathways involved in gastrointestinal
35 motility. They may be also used to scan for motilin agonists and antagonists; as in particular to test the specificity of identified ligands.

The following, non-limiting Examples are presented to better illustrate the invention.

5

EXAMPLE 1

Sequence Comparison of MTL-R1 (GPR38) to human GHS-R, Puffer Fish 75E7 and Identification of Alternatively Spliced Forms.

- 10 Inspection of the MTL-1 genomic DNA sequence revealed two potential mRNA splice sites corresponding to consensus boundaries for exon/intron junctions. An imperfect donor site (TGC/gt) was found at nucleotides 1929-31 (Fig. 1), a perfect donor site (CCG/gt) was found at nucleotides 2080-82, and a single perfect splice acceptor site (sequence
15 [pyrimidine-rich stretch ag/TG]) was observed at nucleotides 2729-32. To determine which splice forms exist naturally, RACE (rapid amplification of cDNA ends) was performed on thyroid poly (A)+ mRNA and RT-PCR (reverse transcriptase polymerase chain reaction) was conducted on HEK-293/aeq17 cells transfected with the MTL-1
20 genomic DNA construct. Directional RACE reactions were conducted on thyroid poly (A)+ mRNA that had previously been shown by RNase protection assay to contain transcripts for MTL-1R. Primer AP1 5'-CCA TCC TAA TAC GAC TCA CTA TAG GGC-3' (SEQ.ID.NO.:8) corresponds to the 5' end of the coding region including the
25 presumptive Met initiation codon located within the cloning vector. 5'RACE1 corresponds to the 3' end of the MTL-1R coding region including the translation termination codon TAA. 5' RACE1: 5'-TTA TCC CAT CGT CTT CAC GTT AGC GCT TGT CTC-3' (SEQ.ID.NO.:9).
- 30 RACE reactions were carried out on 1 µg of thyroid poly (A)+ mRNA using the Marathon cDNA amplification/advantage PCR kit as per the manufacturer's instructions (Clontech) using the following Touchdown PCR amplification conditions: 94°C for 1 min., 5 cycles of 94°C for 30 sec. and 72°C for 4 min.; 5 cycles of 94°C for 30 sec. and
35 70°C for 4 min.; 25 cycles of 94°C for 20 sec and 68°C for 4 min. An approximately 1.4 kb amplified product was identified which hybridized

- 10 -

with a ^{32}P -labeled probe derived from the TM 2-4 region (3F/4R probe) of the MTL-R. This product was subcloned into PCR-Script vector (Invitrogen) and sequenced.

As diagrammed in Figures 6A-C, DNA sequence analysis revealed two distinct sequences corresponding to alternative use of two splice donor sequences and a common splice acceptor sequence. These results were confirmed by transfecting the MTL-1 genomic construct containing the complete ORF interrupted by a single intron of approximately 0.7 kb into HEK-293/aeq17 cells. mRNA was the isolated (Poly (A)⁺ Pure Kit, Ambion) and shown by Northern blot analysis using the 3F/4R probe to give two hybridizing bands: 2.4 kb containing the unspliced intron and approximately 1.4 kb encoding spliced forms. RT-PCR was then performed (Superscript 2 One-Step Kit, Life Technologies) on MTL-1 mRNA from transfected HEK-293/aeq17 cells using the forward primer 5' RACE1 and reverse primer 3' RACE2 (TM5 region): 5'-CTG CCC TTT CTG TGC CTC AGC ATC CTC TAC-3' (SEQ.ID.NO.:10)

An approximately 500 bp product was cloned (TA vector pCR2.2, Invitrogen), sequenced and shown to be a mixture of both splice forms. Assembly of the complete ORF for MTL-1A without intronic sequence was performed by ligation of an exon 1 fragment (Not I digestion of a MTL-1 plasmid containing the intron in pCDNA-3) to pCDNA-3.1 containing a Not I/EcoRI exon 2 fragment.

To document protein expression, an MTL-1A plasmid encoding a amino-terminal FLAG epitope was constructed by ligation of a Pme I fragment from the MTL-1A/pCDNA-1.1 vector into the EcoRV site of pFLAG/CMV-2 vector (Kodak Imaging Systems). Following transfection of this plasmid into HEK-293/aeq17 cells, a protein of the expected size (approximately 48 kDa) was detected in crude cell membranes by immunoblot analysis.

EXAMPLE 2

Identification of Ligand Specific to Motilin Receptor

To identify a ligand for this orphan GPCR and to determine whether the full length, 7 TM domain GPR38-A is a functional GPCR, a

high-throughput assay was developed which measures Ca^{2+} release with the bioluminescent Ca^{2+} sensitive aequorin reporter protein (capable of measuring ligand-induced IP_3 -coupled mobilization of intracellular calcium and concomitant calcium-induced aequorin bioluminescence).

- 5 Expression of GPR38-A in cell membranes was confirmed using epitope-tagged protein which revealed a single protein species with a molecular weight of approximately 45,000 daltons.

- A broad set of peptide and non-peptide molecules was tested at a single concentration in transiently transfected HEK-293/aeq17 cells (100
10 nM peptide, 10 μM non-peptide). Significant bioluminescent responses (> 4-fold over background) were recorded for the peptide motilin and the non-peptide macrolide erythromycin, which was reported to be a competitive agonist at motilin receptors. Full dose-response curves confirmed this observation. The half-maximal effective concentration
15 (EC_{50}) for human/porcine motilin was 2.1 +/- 0.5 nM motilin whereas erythromycin was considerably less potent (2000 +/- 210 nM; as expected from studies performed on native motilin receptors).

- The signal transduction pathway for the cloned GPR38-A motilin receptor (MTL-R1A) is through activation of phospholipase C, which
20 has been reported for native motilin receptors. Direct radioligand binding studies with [^{125}I] human motilin on cell membranes prepared from transfected cells show that MTL-R1A confers high affinity and specific binding (K_d = 0.1 nM; B_{max} = 240 fmol/mg protein) which are strongly G protein coupled (> 80% inhibition of binding with 100 nM
25 $\text{GTP}\gamma\text{S}$).

EXAMPLE 3

Functional Activation of the MTL-1A Receptor

30

- The aequorin bioluminescence assay is a reliable test for identifying G protein-coupled receptors which couple through the $\text{G}\alpha$ protein subunit family consisting of G_q and G_{11} which leads to the activation of phospholipase C, mobilization of intracellular calcium and
35 activation of protein kinase C. Measurement of MTL-1A expression in the aequorin-expressing stable reporter cell line 293-AEQ17 (Button,

D. et. al., 1993 *Cell Calcium* 14: p. 663-671.) was performed using a Luminoskan RT luminometer (Labsystems Inc., Gaithersburg, MD).

293-AEQ17 cells (8 x 10⁵ cells plated 18 hrs. before transfection in a T75 flask) were transfected with 22 µg of human MTL-R1A plasmid DNA: 264 µg lipofectamine. Following approximately 40 hours of expression the apo-aequorin in the cells was charged for 4 hours with coelenterazine (10 µM) under reducing conditions (300 µM reduced glutathione) in ECB buffer (140 mM NaCl, 20 mM KCl, 20 mM HEPES-NaOH [pH=7.4], 5 mM glucose, 1 mM MgCl₂, 1 mM CaCl₂, 0.1 mg/ml bovine serum albumin). The cells were harvested, washed once in ECB medium and resuspended to 500,000 cells/ml. 100 µl of cell suspension (corresponding to 5x10⁴ cells) was then injected into the test plate, and the integrated light emission was recorded over 30 seconds, in 0.5 second units. 20 µL of lysis buffer (0.1% final Triton X-100 concentration) was then injected and the integrated light emission recorded over 10 seconds, in 0.5 second units. The "fractional response" values for each well were calculated by taking the ratio of the integrated response to the initial challenge to the total integrated luminescence including the Triton X-100 lysis response.

20

EXAMPLE 4

Binding of [¹²⁵I] Human Motilin to Crude Membranes from HEK-293 Cells transfected with the MTL-R1A cDNA.

25 The binding of [¹²⁵I] human motilin to crude membranes prepared from HEK-293/aeq17 cell transfectants was performed as follows. Crude cell membranes were prepared on ice, 48 hrs. post-transfection. Each T-75 flask was washed twice with 10 ml of PBS, once with 1 ml homogenization buffer (50 mM Tris-HCl [pH 7.4], 10 mM MgCl₂. 10 ml of homogenization buffer was added to each flask, cells were removed by scraping and then homogenized using a Polytron device (Brinkmann, Syosset, NY; 3 bursts of 10 sec. at setting 4). The homogenate was centrifuged for 20 min. at 11,000 x g at 0°C and the resulting crude membrane pellet (chiefly containing cell membranes and nuclei) was resuspended in homogenization buffer supplemented with 1.5 % BSA (0.5 ml T75 flask) and kept on ice.

35

Binding reactions were performed at 20°C for 1 hr. in a total volume of 0.5 ml containing: 0.1 ml of membrane suspension (approximately 1 µg protein), 10 µl of ¹²⁵I-human motilin, 10 µl of competing drug and 380-390 µl of homogenization buffer. Bound radioligand was separated by rapid vacuum filtration (Brandel 48-well cell harvester) through GF/C filters pretreated for 1 hr. with 0.5% polyethylenimine. After application of the membrane suspension to the filter, the filters were washed 3 times with 3 ml each of ice-cold 50 mM Tris-HCl [pH 7.4], 10 mM MgCl₂, and the bound radioactivity on the filters was quantitated by gamma counting. Specific binding (> 90% of total) is defined as the difference between total binding and non-specific binding conducted in the presence of 100 nM unlabeled human motilin. Competition binding data were analyzed by a nonlinear curve-fitting program (Prism V, version 2.0; GraphPad Software, San Diego, CA). Results shown are the means (+/- SEM) of triplicate determinations; Human motilin was radiolabeled with ¹²⁵I at 7Tyr to a specific activity of approximately 2000 Ci/mmol (Woods Assay, Portland, OR).

Structure-function analysis suggest that the motilin peptide minimally contains an N-terminal region (amino acids 1-7) essential for activity, linked to a C-terminal alpha helical domain which stabilizes the N-terminal active site region activity. The rank order of potency of several motilin peptide analogs in the MTL1-A functional and binding assays correlates with their reported potency measured by *in vitro* contractility assays (Table 1) performed on native motilin receptors in intestinal tissue. These results are summarized in Table 1 below.

Ligand	Cloned MTL-1A Receptor (human)	
	Aequorin Assay (EC ₅₀ nM)	[¹²⁵ I] hmotilin binding (IC ₅₀ ,nM)
human motilin (MTL)	2.1	0.5
erythromycin	2000	427
roxithromycin	1950	613
metoclopramide	>10,000	>10,000
cisapride	>10,000	>10,000

canine motilin	4.4	0.2
Leu13 MTL	3.9	0.2
1-11 MTL	138	127
1-12 MTL	72	14
1-13 MTL	3.8	0.9
1-19 MTL	4.1	0.3
10-22 MTL	>10,000	1100

The unrelated prokinetic agents metoclopramide and cisapride which have affinity for dopamine and/or 5-HT receptors were inactive, even at high (10 μ M) doses.

5

EXAMPLE 5

Southern Blot Analysis

A genomic Southern blot (EcoRI and BamHI-digested DNA, 10 μ g/lane) was hybridized with the ORF of MTL-1A. Post-hybridizational washing stringencies were at 55°C 4 X SSPE after which the filters were dried and exposed to X-ray film for 5 days at -70°C. Lambda Hind III DNA markers were (in kb), 23.1, 9.4, 6.6, 4.4, 2.3, 2.1. Southern blot analysis conducted in a variety of mammalian and non-mammalian species revealed a simple hybridization pattern consistent with a single, conserved gene encoding MTL-1A.

15

EXAMPLE 6

Puffer Fish Clone 75E7

20

Screening of a puffer fish genomic library identified a single clone (75E7) containing an open reading frame of 363 amino acids with approximately 54% protein sequence identity to the human MTL-R1A. In addition, 75E7 has a similar intron-exon structure to the human MTL-R1A. 75E7 may be the ortholog of the human MTL-R1A.

25

EXAMPLE 7
Expression of the MTL-R1A Gene

5 Transcripts of MTL-1A were detected by RNase Protection Assay (RPA). Synthesis of high-specific activity radiolabeled antisense probes and the RPA was conducted using a kit (MAXIscript and HybSpeed RPA kits; Ambion, Austin, TX) essentially as described by the manufacturer. The anti-sense cRNA MTL-1A probe was
10 synthesized from a cDNA template encompassing nt 1234 to 1516 of the human MTL-1A inserted behind the T7 promoter in pLitmus 28 (New England Biolabs, Beverly, MA). Digestion of the construct with Stu I generated a cRNA transcript approximately 340 nt in size with approximately 60 nt of vector sequence. Input poly A⁺ mRNA
15 (Clontech, Palo Alto, CA) was 5 g for the MTL-1A probe and 0.1 µg for a control human actin probe. Precipitated fragments were subjected to slab-gel electrophoresis (42 cm x 32 cm x 0.4 mm) in 5 % acrylamide/Tris-borate-EDTA buffer containing 8 M urea. The gels were fixed, dried and autoradiographed on film (X-Omat; Kodak,
20 Rochester, NY) for 1-3 days (MTL-1A) or 2 hrs. (actin).

 The distribution profile of MTL-1A mRNA was examined in a panel of GI and non-GI human tissues. MTL-1A mRNA could be detected in whole stomach (most prominently), thyroid, and bone marrow but was absent from several brain regions and other non-CNS
25 tissues.

WHAT IS CLAIMED:

1. A motilin receptor, substantially free from receptor-associated proteins.
- 5 2. A motilin receptor according to Claim 1 which is human.
3. A motilin receptor according to Claim 2 which is
- 10 MTL-R1A having the amino acid sequence SEQ.ID.NO.:3.
4. A motilin receptor according to Claim 3 having the nucleic acid sequence SEQ.ID.NO.:2.
- 15 5. A motilin receptor according to Claim 2 which is MTL-R1B having the amino acid sequence SEQ.ID.NO.:5.
6. A motilin receptor according to Claim 5 having the nucleic acid sequence SEQ.ID.NO.:4.
- 20 7. A motilin receptor according to Claim 6 which is 75E7 having the amino acid sequence SEQ.ID.NO.:7.
8. A method for determining whether a ligand is
- 25 capable of binding to a motilin receptor comprising:
 - (a) transfecting test cells with an expression vector encoding motilin receptor;
 - (b) exposing the test cells to the ligand;
 - (c) measuring the amount of binding of the ligand
 - 30 to the motilin receptor;
 - (d) comparing the amount of binding of the ligand to the motilin receptor in the test cells with the amount of binding of the ligand to control cells that have not been transfected with the motilin receptor

where if the amount of binding of the ligand to the test cells is greater than the amount of binding of the ligand to the control cells, then the substance is capable of binding to motilin receptor.

1/13

TTGAAATTATCTGGTCACTGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTCTGA
GGCGGGTGGACCACCTGGGGTCAGGAGTTCGAGACCAGGCTGGCCAACATGGCGAAACCCTGACTACA
CAAAAAACACAAAATTTAGCCGGGGCTTGGGCGCTCCTGTGCTCCCAGCTACTCAGGAGGCTGAGGTG
GGAGGACTGCTTGAGCCTGGGAGGTGAGGCTGCAGTGAGCTGTGATCGCGCCACTTAACTCCAGCC
TGGACGACAGTGAGACCCTGTCTCAAGAAGAAAAAAGAAAGAAAGAAAAAAGAAAAAAGAA
AATTATTTGGTCAATTATATGGTCAGCTCCCTCCACCACTCGCGAATTTACAGAAGAGGAGAACTGGG
CTGGGCGAGACCAGGACTAGCCCAAGATTACACAAGTTACTCGGTTGTAGAGCCAGGATTAGACAGGA
GAGGCTCTAGATTCTGGTCTAGACTCCCTCCTATTATTTAGCATTATGGCTTCTGAGGATTACCAT
GAGCCCTCCTCCACCGTCAAGCGGCAGCTACCAGCCACCAGACCAGATCCCTTCGAAGGTGCCCGGAG
TACCAGACTGACAAAAGCGCCCGTACAGTGCTCAGTCCTGTAACCAAAGCTGTCTAGGGTGACACAT
CGCTCACCGGACCGGGTAGGGCTCGTGCGCTAAGGGCGCCGGGTATTCCAGTTAGTGAGAGGGAAGC
GCCCTGGAAGTGCATGGGCGGGAGAGGGCGCGGGAGCGGAGCATGGCCGGGCGGGGCGGGCGCG
GCCGTGGGCGGAGACTGCGCGCAGCTAGCTCGGGAGCGCCTCGGAGCC QCCCCGAGAGCCGCTTCT
CGCGCCCCGCGAGCGCAGCGCAGCGCTCCGCCGTCTGACCTGCCGCGCCCCGAGCGTGCGGGCTGGGAA
AGGAGGCGCTCACCGAGAGGGACACGCGCCAGGCTCCCAGCCCCGACCCGGGACGCGGCGGGCGCGCG
GAGCACCCATGGGCAGCCCCCTGGAACGGCAGCGACGGCCCCGAGGGGGCGCGGGAGCCGCCGTGGCCC
GCGCTGCCGCCTTGCGACGAGCGCCGCTGCTCGCCCTTTCCCCTGGGGGCGCTGGTGCCGGTGACCGC
TGTGTGCCTGTGCCTGTTCTGTCGTCGGGGTGAGCGGCAACGTGGTGACCGTGATGCTGATCGGGCGCT
ACCGGGACATGCGGACCACCACCAACTTGTACCTGGGCAGCATGGCCGTGTCCGACCTACTCATCCTG
CTCGGGCTGCCGTTGACCTGTACCGCCTCTGGCGCTCGCGGCCCTGGGTGTTGCGGCCGCTGCTCTG
CCGCTGTCCCTCTACGTGGGCGAGGGCTGCACCTACGCCACGCTGCTGCACATGACCGCGCTCAGCG
TCGAGCGCTACCTGGCCATCTGCCGCCCGCTCCGCGCCCGCTCTGGTCAACCGGCGCCGCGTCCGC
GCGCTCATCGCTGTGCTCTGGGCCGTGGCGTGCTCTCTGCCGGTCCCTTCTTGTTCCTGGTGCGGT
CGAGCAGGACCCCGGCATCTCCGTAGTCCCGGGCCTCAATGGCACCAGCGCGGATCGCCTCCTCGCCTC
TCGCCTCGTCGCCGCCTCTCTGGCTCTCGCGGGCGCCACCGCCGTCCCCGCCGTGGGGCCCCGAGACC
GCGGAGGCCGCGGCGCTGTTACGCCGCAATGCCGGCCGAGCCCCGCGCAGCTGGGCGCGCTGCGTGT
CATGCTGTGGGTCAACACCGCCTACTTCTTCTGCCCTTTCTGTGCCTCAGCATCCTCTACGGGCTCA
TCGGGCGGGAGCTGTGGAGCAGCCGGCGGCCGCTGCGAGGCCCGGCCGCTCGGGGCGGGAGAGAGGC
CACCGGCAGACCGTCCGCGTCTGCGtaagtggagccgcccgtggttccaaagacgcctgcctgcagtc
cgccccgcggggaccgcgcaaacgctccctcccttcccctgctcgcccagctctgggcgccgcttc
cagctcccttccatttgcattccagcctccacccgcccgtcattcccattccccgagaaaaccatgt
cctgtccccccaggagctctgggggaccccagggcgctttgaggggtgggatccccggatccgattcagt
aaccagcagtgcttttccagagcctctgagaccagaaaggagagttggtaattcttaaccaaccacc
tgtagatgccacaaatgaggagtcctcacagtgctcttgagaagacgaggagatttcattaagcta
aaattttttatattaatgttaagtgtgtaaggctaaagtaaaccttgctcgtatcaaaaagtaag
attgtgcagacctgtttagaattcttttcaacagagaacagaaaacttgctcgcgaagtgggtttgt
ggaaggaagcctgccaaaggcggtgtttagagaaattgctccttctggtttatgtccagccttgata
acacatatgggagcctactatgcagttttaaagcaagtatccatgcagcctgcagcctggtcattttt
tctgggggtgaggatctgectaggtagaagttttctctaattttatgtgttacttggtattgcaga
tggttccttgctcggggtgggggtttatttgcctcccaatgcttttggttaatcccgggtgctgtgtctt
atgttgtagTGGTGGTGGTCTGGCATTATAATTTGCTGGTTGCCCTTCCACGTTGGCAGAATCATT
TACATAAACACGGAAGATTCGCGGATGATGTACTTCTCTCAGTACTTTAACATCGTCGCTCTGCAACT
TTTCTATCTGAGCGCATCTATCAACCCAATCCTCTACAACCTCATTTCAAAGAAGTACAGAGCGGCGG
CCTTTAACTGCTGCTCGCAAGGAAGTCCAGGCCGAGAGGCTTCCACAGAAGCAGGGACACTGCGGGG
GAAGTTGCAGGGGACACTGGAGGAGACACGGTGGGCTACACCGAGACAAGCGCTAACGTGAAGACGAT
GGGATAA

FIG. 1

2/13

ATGGGCAGCCCCTGGAACGGCAGCGACGGCCCCGAGGGGGCGCGGGAGCCGCCGTGGCCCGCGCTG
CCGCCTTGCGACGAGCGCCGCTGCTCGCCCTTTCCCTGGGGGCGCTGGTGCCGGTGACCGCTGTG
TGCCTGTGCCTGTTTCGTCTCGGGGTGAGCGGCAACGTGGTGACCGTGATGCTGATCGGGCGCTAC
CGGGACATGCGGACCACCACCAACTTGTACCTGGGCAGCATGGCCGTGTCCGACCTACTCATCCTG
CTCGGGCTGCCGTTTCGACCTGTACCGCCTCTGGCGCTCGCGGCCCTGGGTGTTTCGGGCCGCTGCTC
TGCCGCCTGTCCCTCTACGTGGGCGAGGGCTGCACCTACGCCACGCTGCTGCACATGACCGCGCTC
AGCGTCGAGCGCTACCTGGCCATCTGCCGCCCGCTCCGCGCCCGCTCTTGGTCACCCGGCGCCGC
GTCCGCGCGCTCATCGCTGTGCTCTGGGCCGTGGCGCTGCTCTCTGCCGGTCCCTTCTTGTTCCCTG
GTGGGCGTCGAGCAGGACCCCGGCATCTCCGTAGTCCCGGGCCTCAATGGCACCGCGCGGATCGCC
TCCTCGCCTCTCGCCTCGTCGCCGCCTCTCTGGCTCTCGCGGGCGCCACCGCCGTCCCCGCCGTG
GGGCCCCGAGACCGCGGAGGCCGCGGCGCTGTTTCAGCCGCGAATGCCGGCCGAGCCCCGCGCAGCTG
GGCGCGCTGCGTGTCTGCTGTGGGTACCAACCGCCTACTTCTTCTGCCCTTTCTGTGCCTCAGC
ATCCTCTACGGGCTCATCGGGCGGGAGCTGTGGAGCAGCCGGCGGCCGCTGCGAGGCCCGGCCGCC
TCGGGGCGGGAGAGAGAGGCCACCGGCAGACCGTCCGCGTCTGCTGGTGGTGGTTCTGGCATTATA
ATTTGCTGGTTGCCCTTCCACGTTGGCAGAATCATTTACATAAACACGGAAGATTCGCGGATGATG
TACTTCTCTCAGTACTTTAACATCGTCGCTCTGCAACTTTTCTATCTGAGCGCATCTATCAACCCA
ATCCTCTACAACCTCATTTCAAAGAAGTACAGAGCGGCGGCCCTTTAACTGCTGCTCGCAAGGAAG
TCCAGGCCGAGAGGCTTCCACAGAAGCAGGGACACTGCGGGGGAAGTTGCAGGGGACACTGGAGGA
GACACGGTGGGCTACACCGAGACAAGCGCTAACGTGAAGACGATGGGATAA

FIG.2

3/13

MGSPWNGSDGPEGAREPPWPALPPCDERRCSPFPLGALVPVTAVCLCLFVVGVSGNVVTVMLIGRY
RDMRTTTNLYLGSMVSDLLILLGLPFDLYRLWRSRPWVFGPLLCLSLYVGEGCTYATLLHMTAL
SVERYLAICRPLRARVLVTRRRVRALIAVLWAVALLSAGPFLFLVGVEQDPGISVVPGLNGTARIA
SSPLASSPPLWLSRAPPPSPPSGPETAEEAALFSRECRPSPAQLGALRVMLWVTTAYFFLPFLCLS
ILYGLIGRELWSSRRPLRGPAASGRERGHRQTVRVLLVVVLAFIICWLPFHVGRIIYINTEDSRMM
YFSQYFNIVALQLFYLSASINPILYNLISKXYRAAAFKLLLARKSRPRGFHRSRDTAGEVAGDTGG
DTVGYTETSANVKTMG

FIG.3

4/13

ATGGGCAGCCCCTGGAACGGCAGCGACGGCCCCGAGGGGGCGCGGGAGCCGCGGTGGCCCGCGCTG
CCGCCTTGCGACGAGCGCCGCTGCTCGCCCTTTCCCTGGG&GCGCTGGTGCCGGTGACCGCTGTG
TGCCTGTGCCTGTTTCGTGCTCGGGGTGAGCGGCAACGTGGTGACCGTGATGCTGATCGGGCGCTAC
CGGGACATGCGGACCACCACCAACTTGTACCTGGGCAGCATGGCCGTGTCCGACCTACTCATCCTG
CTCGGGCTGCCGTTTCGACCTGTACCGCCTCTGGCGCTCGCGGCCCTGGGTGTTTCGGGCCGCTGCTC
TGCCGCTGTCCCTCTACGTGGGCGAGGGCTGCACCTACGCCACGCTGCTGCACATGACCGCGCTC
AGCGTCGAGCGCTACCTGGCCATCTGCCGCCCGCTCCGCGCCCGCTCTTGGTACCCGGCGCCGC
GTCCGCGCGCTCATCGCTGTGCTCTGGGCCGTGGCGCTGCTCTCTGCCGGTCCCTTCTTGTTCCCTG
GTGGGCGTCGAGCAGGACCCCGGCATCTCCGTAGTCCCGGGCCTCAATGGCACCAGCGCGGATCGCC
TCCTCGCCTCTCGCCTCGTCGCCGCTCTCTGGCTCTCGCGGGCGCCACCGCCGTCCCCGCCGTG
GGGCCCCGAGACCGCGGAGGCCGCGCGCTGTTTCAGCCGCGAATGCCGGCCGAGCCCCGCGCAGCTG
GGCGCGCTGCGTGTGCTGTGGGTACACCACCGCCTACTTCTTCTGCCCTTTCTGTGCCTCAGC
ATCCTCTACGGGCTCATCGGGCGGGAGCTGTGGAGCAGCCGGCGGCCGCTGCGAGGCCCGGCCGCC
TCGGGGCGGGAGAGAGGCCACCGGCAGACCGTCCGCGTCTGCGTAAGTGGAGCCGCCGTGGTTCC
AAAGACGCCTGCCTGCAGTCCGCCCCGCCGGGACCGCGCAAACGCTGGGTCCCCTTCCCCTGCTC
GCCCAGCTCTGGGCGCCGCTTCCAGCTCCCTTTCCTATTTTCGATTCCAGCCTCCACCCGCCGTGGT
GGTGGTTCTGGCATTATAATTTGCTGGTTGCCCTTCCAGTTGGCAGAATCATTTACATAAACAC
GGAAGATTCGCGGATGATGTACTTCTCTCAGTACTTTAACATCGTCGCTCTGCAACTTTTCTATCT
GAGCGCATCTATCAACCAATCCTCTACAACCTCATTTCAAAGAAGTACAGAGCGGCGGCCTTTAA
ACTGCTGCTCGCAAGGAAGTCCAGGCCGAGAGGCTTCCACAGAAGCAGGGACACTGCGGGGGAAGT
TGCAGGGGACACTGGAGGAGACCGGTGGGCTACACCGAGACAAGCGCTAACGTGAAGACGATGGG
ATAA

FIG.4

5/13

MGSPWNGSDGPEGAREPPWPALPPCDERRCSPFPLGALVPVTAVCLCLFVVGVSIGNVIVMLIGRY
RDMRTTTNLYLGSMVSDLLILLGLPFDLYRLWRSRPWVFGPLLCLSLYVGEGCTYATLLHMTAL
SVERYLAICRPLRARVLVTRRRVRALIAVLWAVALLSAGPFLFLVGVEQDPGISVVPGLNGTARIA
SSPLASSPPLWLSRAPPPSPPSGPETAEEAALFSRECRPSPAQLGALRVMLWVTTAYFFLPFLCLS
ILYGLIGRELWSSRRPLRGPAASGREGRGHROTVRVLRKWSRRGSKDACLSAPPGTAQTLGPLPLL
AQLWAPLPAPFPISIPASTRRGGGSGIYNLLVALPRWQNLHKKHGRFADDVLLSVL

FIG.5

ATG	GGC	AGC	CCC	TGG	AAC	GGC	AGC	GAC	GGC	CCC	GAG	GGG	GGC	CGG	GAG	CGG	CCG	TGG	CCC	GGC	CTG	CCG	CCT	TGC
1 M	G	S	P	W	N	G	S	D	G	P	E	G	A	R	E	P	P	W	P	A	L	P	P	C
GAC	GAG	CGC	CGC	TGC	TCG	CCC	TTT	CCC	CTG	GGG	CGG	CTG	GTG	CCG	GTG	ACC	GCT	GTG	TGC	CTG	TGC	CTG	TTC	
D	E	R	R	C	S	P	F	P	L	G	A	L	V	P	V	T	A	V	C	L	C	L	F	
GTC	GGG	GTG	AGC	GGC	AAC	GTG	GTG	ACC	GTG	ATG	CTG	ATC	GGG	CGC	TAC	CGG	GAC	ATG	CGG	ACC	ACC	ACC	AAC	
V	G	V	S	G	N	V	V	T	V	M	L	I	G	R	Y	R	D	M	R	T	T	T	N	
TAC	CTG	GGC	AGC	ATG	GCC	GTG	TCC	GAC	CTA	CTC	ATC	CTG	CTC	GGG	CTG	GAC	TTC	GAC	CTG	TAC	CGC	CTC	TGG	
Y	L	G	S	M	A	V	S	D	L	L	I	L	L	G	L	P	F	D	L	Y	R	L	W	
TGG	CGG	CCC	TGG	GTG	TTC	GGG	CCG	CTG	CTC	TGC	CGC	CTG	TCC	CTC	TAC	GTG	GGC	GAG	GGC	TGC	ACC	TAC	GCC	
S	R	P	W	V	F	G	P	L	L	C	R	L	S	L	Y	V	G	E	G	C	T	Y	A	
CTG	CTG	CAC	ATG	ACC	GCG	CTC	AGC	GTC	GAG	CGC	TAC	CTG	GCC	ATC	TGC	CGC	CCG	CTC	CGC	GCC	CGC	GTG	TTG	
L	L	H	M	T	A	L	S	V	E	R	Y	L	A	I	C	R	P	L	R	A	R	V	L	
ACC	CGG	CGC	CGC	GTC	CGC	CGC	CTC	ATC	GCT	GTG	CTC	TGG	GCC	GTG	GCG	CTG	CTC	TCT	GCC	GGT	CCC	TTG	TTG	
T	R	R	R	V	R	A	L	I	A	V	L	W	A	V	A	L	L	S	A	G	P	F	L	
CTG	GTG	GGC	GTC	GAG	CAG	GAC	CCC	GGC	ATC	TCC	GTA	GTC	CCG	GGC	CTC	AAT	GGC	ACC	GCG	CGG	ATC	GCC	TCC	
L	V	G	V	E	Q	D	P	G	I	S	V	V	P	G	L	N	G	T	A	R	I	A	S	
CCT	CTC	GCC	TCG	TCG	COG	CCT	CTC	TGG	CTC	TCG	CGG	CGC	CCA	CCG	CCG	TCC	CCG	CCG	TCG	GGG	CCC	GAG	ACC	
P	L	A	S	S	P	P	L	W	L	S	R	A	P	P	P	S	P	P	S	G	P	E	T	
GAG	GCC	GCG	GCG	CTG	TTC	AGC	CGC	GAA	TGC	CGG	CCG	AGC	CCC	GCG	CAG	CTG	GGC	GCG	CTG	CGT	CTG	ATG	CTG	
E	A	A	A	L	F	S	R	E	C	R	P	S	P	A	Q	L	G	A	L	R	V	M	L	
GTC	ACC	ACC	GCC	TAC	TTC	TTC	CTG	CCC	TTT	CTG	TGC	CTC	AGC	ATC	CTC	TAC	GGG	CTC	ATC	GGG	CGG	GAG	CTG	
V	T	T	A																					

FIG. 6A

7/13

(Donor A)

CgtAAGTGGAGCCGCGTGGTTCCAAAGAGCCCTGCCTGCAGTCCGCCCGCGGGGACCGGCAAGCGTGGTCCCT
TCCCCTGCTGCCCCAGCTCTGGGCGCGCTCCAGCTCCCTTCCCTATTCGATTCCAGCTCCACCGCCGgt...+569 bp
(Donor B)

FM-1A: 7TM, 403 amino acids

IM6

ag/CTG GTG GTT CTG GCA TTT ATA ATT TGC TGG TTC CCC TTC CAC GTT GGC AGA ATC
L V V V L A F I I C W L P F H V O R I

IMZ

ATT TAC ATA AAC ACG GAA GAT TCG CGG ATG ATG TAC TTC TCT CAG TAC TTT AAC ATC GTC GCT CTG CAA CTT TTC
I Y I N T E D S R M M Y F S Q Y F N I V A L Q L F

TAT CTG AGC GCA TCT ATC AAC CCA ATC CTC TAC AAC CTC ATT TCA AAG AAG TAC AGA GCG GCG GCC TTT AAA CTG
Y L S A S I N P I L Y N L I S K K Y R A A F K L

CTG CTC GCA AGG AAG TCC AGG CCG AGA GGC TTC CAC AGA AGC AGG GAC ACT GCG GGG GAA GTT GCA GGG GAC ACT
L L A R K S R P R G F H R S I R D T A G E V A G D T

GGA GGA GAC ACG GTG GGC TAC ACC GAG ACA AGC GCT AAC GTG AAG ACG ATG GGA TAA
G G D T V G Y T E T S A N V K T M G *

403

FIG.6B

8/13

FM-1B: 5TM, 387 amino acids

CGT AAG TGG AGC CGC CGT GGT TCC AAA GAC GCC TGC CTG CAG TCC GCC CCG GGG ACC GCG CAA ACG CTG
 R K W S R R G S K D A C L Q S A P P G T A Q T L

GGT CCC CTT CCC CTG CTC GCC CAG CTC TGG GCG CCG CTT CCA GCT CCC TTT CCT ATT TCG ATT CCA GCC TCC ACC
 G P L P L A Q L W A P L P A P F P I S I P A S T

CGC CGT GGT GGT TCT GGC ATT TAT AAT TTG CTG GTT GCC CTT CCA CGT TGG CAG AAT CAT TTA CAT AAA CAC
 R R G G S G I Y N L L V A L P R W Q N H L H K H

GGA AGA TTC GCG GAT GAT GTA CTT CTC TCA GTA CTT TAA
 G R F A D D V L L S V L *

387

FIG.6C

9/13

ATGCCCTGGACCAGACCCAGGTGGACCTCCATGCTGCTGCAGCAGAGACCATGGACCAGTACACC
ACGGACGACCACCCTACGAGGGCTCCCTCTTCCCCGCGTCCACCCTCATCCCCGTACGGTCATC
TGCATCCTCATCTTCGTGGTCGGCGTGACCGGCAACACCATGACCATCCTCATCATCCAGTACTTC
AAGGACATGAAGACCACCACCAACCTGTACCTGTCCAGCATGGCCGTGTCCGACCTCGTCATCTTC
CTCTGCCCTGCCCTTCGACCTGTACCGCCTGTGGAAGTACGTGCCGTGGCTGTTGCGCGAGGCCGTG
TGCCGCCTCTACCACTACATCTTCGAAGGCTGCACGTGCGCCACCATCCTCCACATCACGGCCCTG
AGCATCGAGCGCTACCTGGCCATCAGCTTCCCCCTCAGGAGCAAGGTGATGGTGACCAGGAGAAGG
GTCCAGTACATCATCCTGGCCCTGTGGTGCTTCGCCCTGGTGTCGGCCGCTCCCACGCTCTTCCTG
GTCGGGGTGGAGTACGACAACGAGACGCACCCCGACTACAACACGGGCCAGTGCAAGCACACGGGC
TACGCCATCAGCTCGGGGCAGCTGCACATCATGATCTGGGTGTCCACCACCTACTTCTTGCCCG
ATGCTGTGTCTCCTCTTCCTCTACGGCTCCATCGGGTGCAAGCTGTGGAAGAGCAAGAACGACCTG
CAGGGCCCGTGCGCCCTGGCCCGGAGAGGTGCGACAGGCAAACGGTGAAGATCCTGGTGGTGGTG
GTGCTGGCCTTCATCATCTGCTGGCTGCCCTACCACATCGGCAGGAACCTGTTGCCCAGGTGGAC
GACTACGACACGGCCATGCTCAGCCAGAATTTCAACATGGCCTCCATGGTGCTCTGCTACCTCAGC
GCCTCCATCAACCCCGTCGTCTACAACCTGATGTCGAGGAAGTACCGGGCCGCCGCAAGCGCCTC
TTCCTGCTCCACCAGAGACCCAAGCCGGCCACCGGGGGCAGGGGCAGTTTTGCATGATCGGCCAC
AGCCCCACCCTGGACGAGAGCCTGACGGGGGTGTGA

FIG. 7

10/13

MPWTRPQVDLHAAAAETMDQYTTDDHHYEGSLFPASTLIPVTVICILIF W GVTGNT
MTILIIQYFKDMKTTTNLYLSSMAVSDLVIFLCLPFDLYRLWKYVPWLFGEAVCRLY
HYIFEGCTSATILHITALSIERYLAISFPLRSKVMVTRRRVQYIILALWCFALVSAA
PTLFLVGVEYDNETHPDYNTGQCKHTGYAISSGQLHIMI WVSTTYFFCPMLCLFLY
GSIGCKLWKSNDLQGPCALARERSHRQTVKILVVVVLAFFIICWLPHYHGRNLFAQV
DDYDTAMLSQNFNMAVMVLCYLSASINPVVYNLMSRKYRAAAKRLFLLHQRPKPAHR
GQGQFCMIGHSPTLDESLTGV

FIG.8

11/13

pu75E7 1 ..MPWTRPQVDLHAAAAETMDQYTTDDHHYEGSLFPASTLIPVTVICILI 48
 || | || || |:||||':|: :
 huMTLR 1 MGSPWNGS..DGPEGAREPPWPALPPCDERRCSPFPLGALVPVTAVCLCL 48

 49 FVVGVTGNTMTILIIQYFKDMKTTTNLYLSSMAVSDLVIFLCCLPFDLYRL 98
 |||||.|| .|:|:| | :|:|:||||||| |||||.| |||||
 49 FVVGVSIGNVVTVMILIGRYRDMRTTNLYLGSMAVSDLLILLGLPFDLYRL 98

 99 WKYVPWLFGEAVCRLYHYIFEGCTSATILHITALSIERYLAISFPLRSKV 148
 |: ||.|| .||| |: |||| ||:||.||||:||||| |||. :|
 99 WRSRPWVFGPLLCLSLYVGEGETYATLLHMTALSVERYLAICRPLRARV 148

 149 MVTRRRVQYIILALWCFALVSAAPTLFLVGVEYD..... 182
 :|||||. :| || ||.|| | ||||| |
 149 LVTRRRVRALIAVLWAVALLSAGPFLFLVGVEQDPGISVVPGLNGTARIA 198

 183NETHPDYNTGQCKHTGYAISS.....GQLHIM 209
 | .| | :| | :|
 199 SSPLASSPPLWLSRAPPPSPSPGPETAEEAALFSRECRPSPAQLGALRVM 248

 210 IWVSTTYFFCPMLCLFLYGSIGCKLWKSNDLQGPCALARERSHRQTVK 259
 :|.| ||| | ||| ||| || :|| |: |.|| | ||| ||||| :
 249 LWVTTAYFFLPFLCLSILYGLIGRELWSSRRPLRGPAASGRERGHRTVR 298

 260 ILVVVVLAFAICWLPYHIGRNLFQVDDYDTAMLSQNFNMASMLCYLSA 309
 :|.||||||| |||||:|:| :| :| || ||. :| |||||
 299 VLLVVVLAFIICWLPFHVGRIIYINTEDSRMMYFSQYFNIVALQLFYLSA 348

 310 SINPVVYNIMSRRYRAAAKRLFLLHQ.RPKPAHRGQ...GQFCMIGHSP 355
 |||||.|||. |:||||| :| | . ||: || . |: |
 349 SINPILYNLISKKYRAAAFKLLLARKSRPRGFHRSRDTAGEVAGDTGGDT 398

 356 LDESLTGV..... 363
 |
 399 VGYTETSANVKTMG 412

FIG.9

12/13

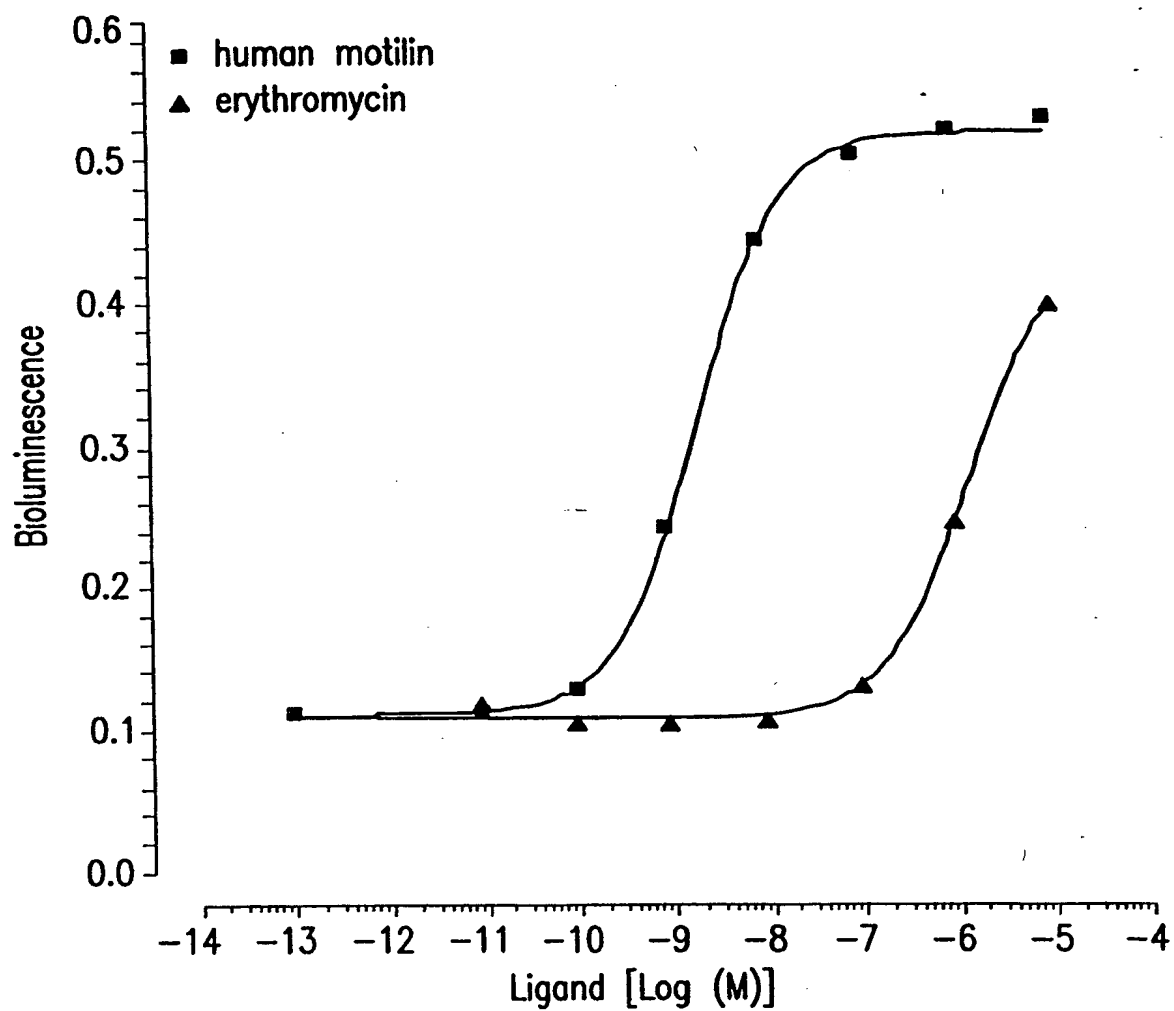


FIG.10

13/13

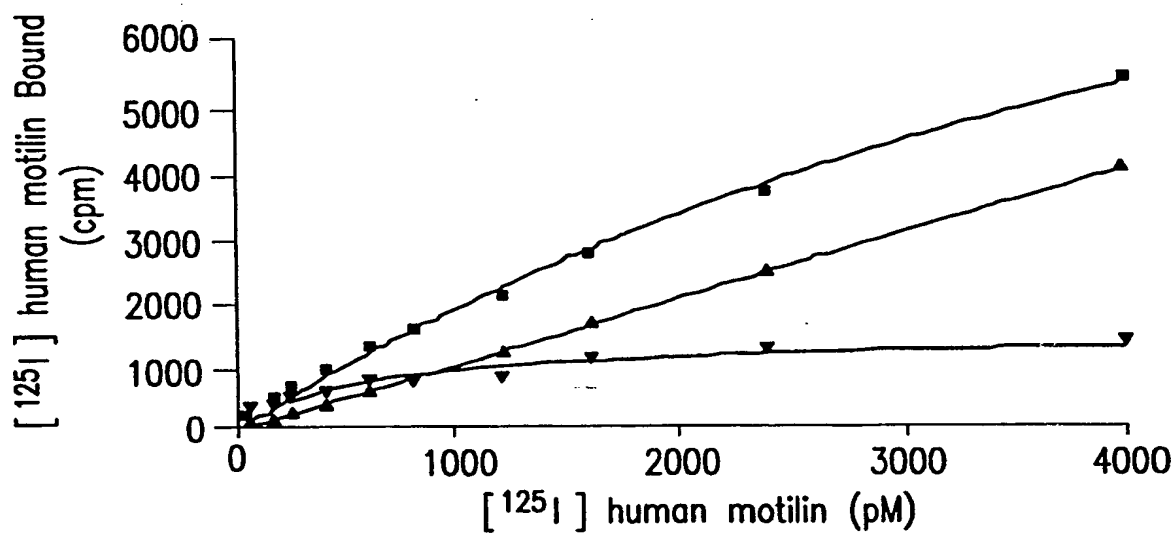


FIG. 11

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Merck & Co., Inc.
- (ii) TITLE OF INVENTION: CLONING AND IDENTIFICATION
OF THE MOTILIN RECEPTOR
- (iii) NUMBER OF SEQUENCES: 15
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Merck & Co., Inc.
 - (B) STREET: P.O. Box 2000, 126 E. Lincoln Ave.
 - (C) CITY: Rahway
 - (D) STATE: NJ
 - (E) COUNTRY: USA
 - (F) ZIP: 07065-0900
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: Windows
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0b
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 60/089,098
 - (B) FILING DATE: 12-JUN-1998
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Giesser, Joanne M
 - (B) REGISTRATION NUMBER: 32,838
 - (C) REFERENCE/DOCKET NUMBER: 20251 PCT
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 732-594-3046
 - (B) TELEFAX: 732-594-4720
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3066 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGAAATTAT	CTGGTCACTG	CCGGGCGCGG	TGGCTCACGC	CTGTAATCCC	AGCACTTTGG	60
GAGGTCGAGG	CGGGTGGACC	ACCTGGGGTC	AGGAGTTCGA	GACCAGGCTG	GCCAACATGG	120
CGAAACCCTG	ACTACACAAA	AAACACAAAA	TTTAGCCGGG	GCTTGGGGCG	TCCTGTGCTC	180
CCAGCTACTC	AGGAGGCTGA	GGTGGGAGGA	CTGCTTGAGC	CTGGGAGGTC	GAGGCTGCAG	240
TGAGCTGTGA	TCGCGCCACT	TAAACTCCAG	CTGGACGAC	AGTGAGACCC	TGCTCAAGA	300
AGAAAAAAG	AAAGAAAGAA	AGAAAAAAG	AAAAAAAAGA	AATTTATTTGG	TCAATTATAT	360
GGTCAGCTCC	CTCCACCAC	CGCGAATTTA	CAGAAGAGGA	GAAC TGGGCT	GGGCGAGACC	420
AGGACTAGCC	CAAGATTACA	CAAGTTACTC	GGTTGTAGAG	CCAGGATTAG	ACAGGAGAGG	480
CTCTAGATT	TGGTCTAGAC	TCCCCTCC	TTATTTAGCA	TTATGGCTTC	CTGAGGATTA	540
CCATGAGCCC	TCC TCCACC	TCAAGCGGCA	GC TACCAGCC	ACCAGACCAG	ATCCCTTCGA	600
AGGTGCCCCG	AGTACCAGAC	TGACAAAAGC	GCCCCGTACAG	TGCTCAGTCC	TGTAACCAAA	660
GCTGTCTAGG	GTGCAGACAT	CGCTCACCGG	ACCGGGTAGG	GCTCGTGCGC	TAAGGGCGCC	720
GGGTATTCCA	GTTAGTGGAG	AGGGAAGCGC	CC TGGAACTG	CATGGGCCCC	GGAGAGGGCG	780
CGGGAGCGGA	GCA TGGCCG	GCCGGGGCGG	GCCGCGGCCG	TGGGCGGAGA	CTGCGCGCAG	840
CTAGCTCGGG	AGCGCCTCGG	AGCCCACCCC	GCAGAGCCGC	TTCTCGCGCC	CCGCAGCGCA	900
GCGCAGCGCT	CCGCCGTCTG	ACCTGCCCGC	CCC GCAGCGT	CGGGGCTGGG	AAAGGAGGCG	960
CTCACCGAGA	GGGACCACGC	GCCAGGCTCC	CAGCCCGACC	CGGGACGCGG	CGGCCGCGCG	1020
GAGCACCCAT	GGGCAGCCCC	TGGAACGGCA	GCGACGGCCC	CGAGGGGGCG	CGGGAGCCGC	1080
CGTGGCCCCG	GCTGCCGCC	TGCGACGAGC	GCCGCTGCTC	GCCCTTTCCC	CTGGGGGCGC	1140
TGGTGCCGGT	GACCGCTGTG	TGCCTGTGCC	TGTTCTGCTG	CGGGGTGAGC	GGCAACGTGG	1200
TGACCGTGAT	GCTGATCGGG	CGTACC	ACATGCGGAC	CACCACCAAC	TTGTACCTGG	1260
GCAGCATGGC	CGTGTCCGAC	CTACTCATCC	TGCTCGGGCT	CCCGTTTCGAC	CTGTACCGCC	1320
TCTGGCGCTC	GCGGCCCTGG	GTGTTCCGGC	CGCTGCTCTG	CCGCCTGTCC	CTCTACGTGG	1380
GCGAGGGCTG	CACCTACGCC	ACGCTGCTGC	ACATGACCGC	GCTCAGCGTC	GAGCGCTACC	1440
TGGCCATCTG	CCGCCCCGTC	CGCGCCCCG	TC TTTGGTCA	CCGGCGCCGC	GTCCGCGCGC	1500
TCATCGCTGT	GCTCTGGGCC	GTGGCGCTGC	TCTCTGCCCG	TCCCTTCTTG	TTCTTGGTGG	1560
TGGTCGAGCA	GGACCCCGGC	ATCTCCGTAG	TCCCGGGCCT	CAATGGCACC	CGCGGATCCG	1620
CCTCCTCGCC	TCTCGCCTCG	TCGCCGCCCTC	TCTGGCTCTC	GCGGGCGCCA	CCGCCGTCCC	1680
CGCCGTCCGG	GCCCGAGACC	GCGGAGGCCG	CGGCGCTGTT	CAGCCGCGAA	TGCCGGCCGA	1740
GCCCCGCGCA	GCTGGGCGCG	CTGCGTGTC	TGCTGTGGGT	CACCACCGCC	TACTTCTTCC	1800
TGCCCTTTCT	GTGCCTCAGC	ATCCTCTACG	GGCTCATCGG	GCGGGAGCTG	TGGAGCAGCC	1860
GGCGCCGCTG	GCGAGGCCG	GCCGCCTCGG	GGCGGGAGAG	AGGCCACCGG	CAGACCGTCC	1920
GCGTCTGCG	TAAGTGAGC	CGCCGTGGTT	CCAAAGACGC	CTGCCTGCAG	TCCGCCCGCC	1980
CGGGGACCGC	GCAAACGCTG	GGTCCCCCTC	CCCTGCTCGC	CCAGCTCTGG	GCGCCGCTTC	2040
CAGCTCCCTC	CTATTTTCGAT	TCCAGCCTCC	ACCCGCCGGT	ACTTCCCATC	CCCCGAGAAA	2100
ACCATGTCC	GTCCCCCAGG	AGCTCTGGGG	GACCCAGGG	CGCTTTGAGG	GTGGGATCCC	2160
CGGATCCGAT	TCAGTAACCA	GCAGTGCTTT	TCCAGAGCCT	CTGAGACCAG	AAAGGAGAGT	2220
TGGTAATTCT	TAATCCAACC	ACCTGTTAGA	TGCCACAAAT	GAGGAGTCC	CACAGTGCTC	2280
TTGAGAAGAC	GAGGGAGATT	TCATTAAGCT	AAAATTTTTT	ATTTAATGTT	AAGTGATGCT	2340
GAAGGCTAAA	GTAAACCTTG	CTCGTATCAA	AAAGTAAAGA	TTGTGCAGAC	CTGTTGTAGA	2400
ATTCTTTTCA	ACAGAGAACA	GAAAACTTGT	CTCCGAAGTG	GGTTTGTGGA	AGGAAGCCTG	2460
CCAAGGCGGC	TTGTTTCAGAG	AAATTGCTCC	TTCTGGTTTA	TGTCCAGCCT	TGATAACACA	2520
TATGGGAGCC	TACTATGCAG	TTTTAAAGCA	AGTATCCATG	CAGCCTGCAG	CTGGTTCATT	2580
TTTTCTGGGG	TGAGGATCTG	CCTAGGTAGA	AGTTTTCTCT	AATTTATTTT	GCTGTTACTT	2640
GTTATTGCAG	ATGGTTCCCT	GTCGGGGTGG	GGGGTTTATT	TGCTTCCCAA	TGCTTTTGTT	2700
AATCCCGGTG	CTGTGTCTTA	TGTTGCAGTG	GTGGTGGTTC	TGGCATTTAT	AATTTGCTGG	2760
TTGCCCTTCC	ACGTGGGCAG	AATCATTTAC	ATAAACACGG	AAGATTCGCG	GATGATGTAC	2820
TTCTCTCAGT	ACTTTAACAT	CGTCGCTCTG	CAACTTTTCT	ATCTGAGCGC	ATCTATCAAC	2880
CCAATCTCT	ACAACCTCAT	TTCAAAGAAG	TACAGAGCGG	CGGCCTTTAA	ACTGCTGCTC	2940
GCAAGGAAGT	CCAGGCCGAG	AGGCTTCCAC	AGAAGCAGGG	ACACTGCGGG	GGAAGTTGCA	3000
GGGGACACTG	GAGGAGACAC	GGTGGGCTAC	ACCGAGACAA	GCGCTAACGT	GAAGACGATG	3060
GGATAA						3066

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1239 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC      60
GCGCTGCCGC CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGC GCTGGTGCCG      120
GTGACCGCTG TGTGCCTGTG CCTGTTCGTC GTCGGGGTGA GCGGCAACGT GGTGACCGTG      180
ATGCTGATCG GCGCTACCG GGACATGCGG ACCACCACCA ACTTGTACCT GGGCAGCATG      240
GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTTC ACCTGTACCG CCTCTGGCGC      300
TCGCGGCCCT GGGTGTTCGG GCCGCTGCTC TGCCGCTGT CCCTCTACGT GGGCGAGGGC      360
TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC      420
TGCCGCCCCG TCCGCGCCCCG CGTCTTGGTC ACCCGCGCC GCGTCCGCGC GCTCATCGCT      480
GTGCTCTGGG CCGTGGCGCT GCTCTCTGCC GGTCCCTTCT TGTTCCTGGT GGGCGTCGAG      540
CAGGACCCCG GCATCTCCGT AGTCCCAGGG CTCAATGGCA CCGCGCGGAT CGCCTCCTCG      600
CCTCTCGCCT CGTCGCGCCG TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCCGCCGTCG      660
GGGCCCCGAG CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG      720
CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCAACG CCTACTTCTT CCTGCCCTTT      780
CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGGAGC TGTGGAGCAG CCGGCGGCCG      840
CTGCGAGGCC CGGCCGCCTC GGGGCGGGAG AGAGGCCACC GGCAGACCGT CCGCTCCTG      900
CTGGTGGTGG TTCTGGCATT TATAATTTGC TGGTTGCCCT TCCACGTTGG CAGAATCATT      960
TACATAAACA CGGAAGATTC GCGGATGATG TACTTCTCTC AGTACTTTAA CATCGTCGCT      1020
CTGCAACTTT TCTATCTGAG CGCATCTATC AACCCAATCC TCTACAACCT CATTTCAAAG      1080
AAGTACAGAG CGGCGGCCTT TAAACTGCTG CTCGCAAGGA AGTCCAGGCC GAGAGGCTTC      1140
CACAGAAGCA GGGACACTGC GGGGGAAGTT GCAGGGGACA CTGGAGGAGA CACGGTGGGC      1200
TACACCGAGA CAAGCGCTAA CGTGAAGACG ATGGGATAA      1239

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1             5             10             15
Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
 20             25             30
Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
 35             40             45
Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
 50             55             60
Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
 65             70             75             80

```

```

Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr
      85      90      95
Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg
      100     105     110
Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His
      115     120     125
Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu
      130     135     140
Arg Ala Arg Val Leu Val Thr Arg Arg Arg Val Arg Ala Leu Ile Ala
      145     150     155     160
Val Leu Trp Ala Val Ala Leu Leu Ser Ala Gly Pro Phe Leu Phe Leu
      165     170     175
Val Gly Val Glu Gln Asp Pro Gly Ile Ser Val Val Pro Gly Leu Asn
      180     185     190
Gly Thr Ala Arg Ile Ala Ser Ser Pro Leu Ala Ser Ser Pro Pro Leu
      195     200     205
Trp Leu Ser Arg Ala Pro Pro Pro Ser Pro Pro Ser Gly Pro Glu Thr
      210     215     220
Ala Glu Ala Ala Ala Leu Phe Ser Arg Glu Cys Arg Pro Ser Pro Ala
      225     230     235     240
Gln Leu Gly Ala Leu Arg Val Met Leu Trp Val Thr Thr Ala Tyr Phe
      245     250     255
Phe Leu Pro Phe Leu Cys Leu Ser Ile Leu Tyr Gly Leu Ile Gly Arg
      260     265     270
Glu Leu Trp Ser Ser Arg Arg Pro Leu Arg Gly Pro Ala Ala Ser Gly
      275     280     285
Arg Glu Arg Gly His Arg Gln Thr Val Arg Val Leu Leu Val Val Val
      290     295     300
Leu Ala Phe Ile Ile Cys Trp Leu Pro Phe His Val Gly Arg Ile Ile
      305     310     315     320
Tyr Ile Asn Thr Glu Asp Ser Arg Met Met Tyr Phe Ser Gln Tyr Phe
      325     330     335
Asn Ile Val Ala Leu Gln Leu Phe Tyr Leu Ser Ala Ser Ile Asn Pro
      340     345     350
Ile Leu Tyr Asn Leu Ile Ser Lys Tyr Arg Ala Ala Ala Phe Lys
      355     360     365
Leu Leu Leu Ala Arg Lys Ser Arg Pro Arg Gly Phe His Arg Ser Arg
      370     375     380
Asp Thr Ala Gly Glu Val Ala Gly Asp Thr Gly Gly Asp Thr Val Gly
      385     390     395     400
Tyr Thr Glu Thr Ser Ala Asn Val Lys Thr Met Gly
      405     410

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1390 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC

60

```

GCGCTGCCGC CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGC GCTGGTGCCG 120
GTGACCGCTG TGTGCCTGTG CCTGTTCGTC GTCGGGGTGA GCGGCAACGT GGTGACCGTG 180
ATGCTGATCG GGCGCTACCG GGACATGCGG ACCACCACCA ACTGTACCT GGGCAGCATG 240
GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTCTG ACCTGTACCG CCTCTGGCGC 300
TCGCGGGCCCT GGGTGTTCGG GCCGCTGCTC TGCCGCCTGT CCCTCTACGT GGGCGAGGGC 360
TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC 420
TGCCGCCCCG TCCGCGCCCCG CGTCTTGGTC ACCCGCGCGC GCGTCCGCGC GCTCATCGCT 480
GTGCTCTGGG CCGTGGCGCT GCTCTCTGCC GGTCCCTTCT TGTTCCTGGT GGGCGTCGAG 540
CAGGACCCCG GCATCTCCGT AGTCCCGGGC CTCAATGGCA CCGCGCGGAT CGCCTCCTCG 600
CCTCTCGCCT CGTCGCGGCC TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCCGCCGTCG 660
GGGCCCCGAGA CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG 720
CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCACCG CCTACTTCTT CCTGCCCTTT 780
CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGGAGC TGTGGAGCAG CCGGCGGGCCG 840
CTGCGAGGCC CGGCCGCCTC GGGCGGGAG AGAGGCCACC GGCAGACCGT CCGCGTCCTG 900
CGTAAGTGGA GCGCCGTGG TTCCAAAGAC GCCTGCCTGC AGTCCGCCCC GCGGGGGACC 960
GCGCAAACGC TGGGTCCCTT TCCCCTGCTC GCCCAGCTCT GGGCGCCGCT TCCAGCTCCC 1020
TTTCCTATTT CGATTCCAGC CTCCACCCGC CGTGGTGGTG GTTCTGGCAT TTATAATTTG 1080
CTGGTTGCC TTCCACGTTG GCAGAATCAT TTACATAAAC ACGGAAGATT CGCGGATGAT 1140
GTACTTCTCT CAGTACTTTA ACATCGTCGC TCTGCAACTT TTCTATCTGA GCGCATCTAT 1200
CAACCCAATC CTCTACAACC TCATTTCAAA GAAGTACAGA GCGGCGGCCT TTAAACTGCT 1260
GCTCGCAAGG AAGTCCAGGC CGAGAGGCTT CCACAGAAGC AGGGACACTG CGGGGGAAGT 1320
TGCAGGGGAC ACTGGAGGAG ACACGGTGGG CTACACCGAG ACAAGCGCTA ACGTGAAGAC 1380
GATGGGATAA 1390

```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 386 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1           5           10           15
Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
          20           25           30
Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
          35           40           45
Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
          50           55           60
Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
          65           70           75           80
Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr
          85           90           95
Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg
          100          105          110
Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His
          115          120          125
Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu
          130          135          140
Arg Ala Arg Val Leu Val Thr Arg Arg Arg Val Arg Ala Leu Ile Ala
          145          150          155          160

```

Val Leu Trp Ala Val Ala Leu Leu Ser Ala Gly Pro Phe Leu Phe Leu
 165 170 175
 Val Gly Val Glu Gln Asp Pro Gly Ile Ser Val Val Pro Gly Leu Asn
 180 185 190
 Gly Thr Ala Arg Ile Ala Ser Ser Pro Leu Ala Ser Ser Pro Pro Leu
 195 200 205
 Trp Leu Ser Arg Ala Pro Pro Ser Pro Pro Ser Gly Pro Glu Thr
 210 215 220
 Ala Glu Ala Ala Ala Leu Phe Ser Arg Glu Cys Arg Pro Ser Pro Ala
 225 230 235 240
 Gln Leu Gly Ala Leu Arg Val Met Leu Trp Val Thr Thr Ala Tyr Phe
 245 250 255
 Phe Leu Pro Phe Leu Cys Leu Ser Ile Leu Tyr Gly Leu Ile Gly Arg
 260 265 270
 Glu Leu Trp Ser Ser Arg Arg Pro Leu Arg Gly Pro Ala Ala Ser Gly
 275 280 285
 Arg Glu Arg Gly His Arg Gln Thr Val Arg Val Leu Arg Lys Trp Ser
 290 295 300
 Arg Arg Gly Ser Lys Asp Ala Cys Leu Gln Ser Ala Pro Pro Gly Thr
 305 310 315 320
 Ala Gln Thr Leu Gly Pro Leu Pro Leu Leu Ala Gln Leu Trp Ala Pro
 325 330 335
 Leu Pro Ala Pro Phe Pro Ile Ser Ile Pro Ala Ser Thr Arg Arg Gly
 340 345 350
 Gly Gly Ser Gly Ile Tyr Asn Leu Leu Val Ala Leu Pro Arg Trp Gln
 355 360 365
 Asn His Leu His Lys His Gly Arg Phe Ala Asp Asp Val Leu Leu Ser
 370 375 380
 Val Leu
 385

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGCCCTGGA	CCAGACCCCA	GGTGGACCTC	CATGCTGCTG	CAGCAGAGAC	CATGGACCAG	60
TACACCACGG	ACGACCACCA	CTACGAGGGC	TCCCTCTTCC	CCGCGTCCAC	CCTCATCCCC	120
GTCACGGTCA	TCTGCATCCT	CATCTTCGTG	GTCGGCGTGA	CCGGCAACAC	CATGACCATC	180
CTCATCATCC	AGTACTTCAA	GGACATGAAG	ACCACCACCA	ACCTGTACCT	GTCCAGCATG	240
GCCGTGTCCG	ACCTCGTCAT	CTTCCTCTGC	CTGCCCTTCG	ACCTGTACCG	CCTGTGGAAG	300
TACGTGCCGT	GGCTGTTCGG	CGAGGCCGTG	TGCCGCCTCT	ACCACTACAT	CTTCGAAGGC	360
TGCACGTCGG	CCACCATCCT	CCACATCACG	GCCCTGAGCA	TCGAGCGCTA	CCTGGCCATC	420
AGCTTCCCCC	TCAGGAGCAA	GGTGATGGTG	ACCAGGAGAA	GGGTCCAGTA	CATCATCCTG	480
GCCCTGTGGT	GCTTCGCCCT	GGTGTGCGCC	GCTCCCACGC	TCTTCCTGGT	CGGGGTGGAG	540
TACGACAACG	AGACGCACCC	CGACTACAAC	ACGGGCCAGT	GCAAGCACAC	GGGCTACGCC	600
ATCAGCTCGG	GGCAGCTGCA	CATCATGATC	TGGGTGTCCA	CCACCTACTT	CTTCTGCCCC	660
ATGCTGTGTC	TCCTCTTCCT	CTACGGCTCC	ATCGGGTGCA	AGCTGTGGAA	GAGCAAGAAC	720
GACCTGCAGG	GCCCCGTGCG	CCTGGCCCCG	GAGAGGTCGC	ACAGGCAAAC	GGTGAAGATC	780

CTGGTGGTGG	TGGTGCTGGC	CTTCATCATC	TGCTGGCTGC	CCTACCACAT	CGGCAGGAAC	840
CTGTTGCCCC	AGGTGGACGA	CTACGACACG	GCCATGCTCA	GCCAGAATTT	CAACATGGCC	900
TCCATGGTGC	TCTGCTACCT	CAGCGCCTCC	ATCAACCCCG	TCGTCTACAA	CCTGATGTCG	960
AGGAAGTACC	GGGCCGCCGC	CAAGCGCCTC	TTCCTGCTCC	ACCAGAGACC	CAAGCCGGCC	1020
CACCGGGGGC	AGGGGCAGTT	TTGCATGATC	GGCCACAGCC	CCACCCTGGA	CGAGAGCCTG	1080
ACGGGGGTGT	GA					1092

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Pro	Trp	Thr	Arg	Pro	Gln	Val	Asp	Leu	His	Ala	Ala	Ala	Ala	Glu
1				5					10					15	
Thr	Met	Asp	Gln	Tyr	Thr	Thr	Asp	Asp	His	His	Tyr	Glu	Gly	Ser	Leu
		20						25					30		
Phe	Pro	Ala	Ser	Thr	Leu	Ile	Pro	Val	Thr	Val	Ile	Cys	Ile	Leu	Ile
		35					40					45			
Phe	Val	Val	Gly	Val	Thr	Gly	Asn	Thr	Met	Thr	Ile	Leu	Ile	Ile	Gln
	50					55					60				
Tyr	Phe	Lys	Asp	Met	Lys	Thr	Thr	Thr	Asn	Leu	Tyr	Leu	Ser	Ser	Met
65					70					75				80	
Ala	Val	Ser	Asp	Leu	Val	Ile	Phe	Leu	Cys	Leu	Pro	Phe	Asp	Leu	Tyr
				85					90					95	
Arg	Leu	Trp	Lys	Tyr	Val	Pro	Trp	Leu	Phe	Gly	Glu	Ala	Val	Cys	Arg
			100					105					110		
Leu	Tyr	His	Tyr	Ile	Phe	Glu	Gly	Cys	Thr	Ser	Ala	Thr	Ile	Leu	His
		115					120					125			
Ile	Thr	Ala	Leu	Ser	Ile	Glu	Arg	Tyr	Leu	Ala	Ile	Ser	Phe	Pro	Leu
	130					135					140				
Arg	Ser	Lys	Val	Met	Val	Thr	Arg	Arg	Arg	Val	Gln	Tyr	Ile	Ile	Leu
145					150					155					160
Ala	Leu	Trp	Cys	Phe	Ala	Leu	Val	Ser	Ala	Ala	Pro	Thr	Leu	Phe	Leu
				165					170					175	
Val	Gly	Val	Glu	Tyr	Asp	Asn	Glu	Thr	His	Pro	Asp	Tyr	Asn	Thr	Gly
			180					185					190		
Gln	Cys	Lys	His	Thr	Gly	Tyr	Ala	Ile	Ser	Ser	Gly	Gln	Leu	His	Ile
		195					200					205			
Met	Ile	Trp	Val	Ser	Thr	Thr	Tyr	Phe	Phe	Cys	Pro	Met	Leu	Cys	Leu
	210					215					220				
Leu	Phe	Leu	Tyr	Gly	Ser	Ile	Gly	Cys	Lys	Leu	Trp	Lys	Ser	Lys	Asn
225					230					235					240
Asp	Leu	Gln	Gly	Pro	Cys	Ala	Leu	Ala	Arg	Glu	Arg	Ser	His	Arg	Gln
				245					250					255	
Thr	Val	Lys	Ile	Leu	Val	Val	Val	Val	Leu	Ala	Phe	Ile	Ile	Cys	Trp
		260						265					270		
Leu	Pro	Tyr	His	Ile	Gly	Arg	Asn	Leu	Phe	Ala	Gln	Val	Asp	Asp	Tyr
		275					280						285		

Asp Thr Ala Met Leu Ser Gln Asn Phe Asn Met Ala Ser Met Val Leu
 290 295 300
 Cys Tyr Leu Ser Ala Ser Ile Asn Pro Val Val Tyr Asn Leu Met Ser
 305 310 315 320
 Arg Lys Tyr Arg Ala Ala Lys Arg Leu Phe Leu Leu His Gln Arg
 325 330 335
 Pro Lys Pro Ala His Arg Gly Gln Gly Gln Phe Cys Met Ile Gly His
 340 345 350
 Ser Pro Thr Leu Asp Glu Ser Leu Thr Gly Val
 355 360

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTATCCCATC GTCTTCACGT TAGCGCTTGT CTC

33

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTGCCCTTTC TGTGCCTCAG CATCCTCTAC

30

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC      60
GCGCTGCCGC CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGC GCTGGTGCCG      120
GTGACCGCTG TGTGCCTGTG CCTGTTTCGT GTCGGGGTGA GCGGCAACGT GGTGACCGTG      180
ATGCTGATCG GCGCTACCG GGACATGCGG ACCACCACCA ACTTGACCT GGGCAGCATG      240
GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTTC ACCTGTACCG CCTCTGGCGC      300
TCGCGGCCCT GGGTGTTCGG GCGCTGCTC TGCCGCCTGT CCCTCTACGT GGGCGAGGGC      360
TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC      420
TGCCGCCCCG TCCGCGCCCC CGTCTTGGTC ACCCGCGCGC GCGTCCGCGC GCTCATCGCT      480
GTGCTCTGGG CCGTGGCGCT GCTCTCTGCC GGTCCCTTCT TGTTCCTGGT GGGCGTCGAG      540
CAGGACCCCG GCATCTCCGT AGTCCCGGGC CTCAATGGCA CCGCGCGGAT CGCCTCCTCG      600
CCTCTCGCCT CGTCGCCGCC TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCGCCGTCG      660
GGGCCCCAGA CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG      720
CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCACCG CCTACTTCTT CCTGCCCTTT      780
CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGGAGC TGTGGAGCAG CCGCGGCGCC      840
CTGCGAGGCC CGGCCGCTC GGGCGGGAG AGAGGCCACC GGCAGACCGT CCGCGTCTCT      900

```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1           5           10           15
Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
      20           25           30
Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
      35           40           45
Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
      50           55           60
Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
      65           70           75           80
Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr
      85           90           95
Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg
      100          105          110
Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His
      115          120          125
Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu
      130          135          140

```

Arg	Ala	Arg	Val	Leu	Val	Thr	Arg	Arg	Arg	Val	Arg	Ala	Leu	Ile	Ala
145					150					155					160
Val	Leu	Trp	Ala	Val	Ala	Leu	Leu	Ser	Ala	Gly	Pro	Phe	Leu	Phe	Leu
			165						170					175	
Val	Gly	Val	Glu	Gln	Asp	Pro	Gly	Ile	Ser	Val	Val	Pro	Gly	Leu	Asn
			180					185					190		
Gly	Thr	Ala	Arg	Ile	Ala	Ser	Ser	Pro	Leu	Ala	Ser	Ser	Pro	Pro	Leu
		195					200					205			
Trp	Leu	Ser	Arg	Ala	Pro	Pro	Pro	Ser	Pro	Pro	Ser	Gly	Pro	Glu	Thr
	210					215					220				
Ala	Glu	Ala	Ala	Ala	Leu	Phe	Ser	Arg	Glu	Cys	Arg	Pro	Ser	Pro	Ala
225					230					235					240
Gln	Leu	Gly	Ala	Leu	Arg	Val	Met	Leu	Trp	Val	Thr	Thr	Ala	Tyr	Phe
				245				250						255	
Phe	Leu	Pro	Phe	Leu	Cys	Leu	Ser	Ile	Leu	Tyr	Gly	Leu	Ile	Gly	Arg
			260				265						270		
Glu	Leu	Trp	Ser	Ser	Arg	Arg	Pro	Leu	Arg	Gly	Pro	Ala	Ala	Ser	Gly
	275						280					285			
Arg	Glu	Arg	Gly	His	Arg	Gln	Thr	Val	Arg	Val	Leu				
	290					295					300				

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 154 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGTAAGTGGG	GCCGCCGTGG	TTCCAAAGAC	GCCTGCCTGC	AGTCCGCCCC	GCCGGGGACC	60
GCGCAAACGC	TGGGTCCCCT	TCCCCTGCTC	GCCCAGCTCT	GGGCGCCGCT	TCCAGCTCCC	120
TTTCCTATTT	CGATTCCAGC	CTCCACCCGC	CGGT			154

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 602 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCTGGTGGT	GGTTCTGGCA	TTTATAATTT	GCTGGTTGCC	CTTCCACGTT	GGCAGAATCA	60
TTTACATAAA	CACGGAAGAT	TCGCGGATGA	TGTACTTCTC	TCAGTACTTT	AACATCGTCG	120
CTCTGCAACT	TTTCTATCTG	AGCGCATCTA	TCAACCCAAT	CCTCTACAAC	CTCATTTCAA	180
AGAAGTACAG	AGCGGCGGCC	TTTAAACTGC	TGCTCGCAAG	GAAGTCCAGG	CCGAGAGGCT	240
TCCACAGAAG	CAGGGACACT	GCGGGGGAAG	TTGCAGGGGA	CACTGGAGGA	GACACGGTGG	300
GCTACACCGA	GACAAGCGCT	AACGTGAAGA	CGATGGGATA	ACGTAAGTGG	AGCCGCCGTG	360
GTTCCAAAGA	CGCCTGCCTG	CAGTCCGCCC	CGCCGGGGAC	CGCGCAAACG	CTGGGTCCCC	420

```

TTCCCCTGCT CGCCCAGCTC TGGGCGCCGC TTCCAGCTCC CTTTCCTATT TCGATTCCAG 480
CCTCCACCCG CCGTGGTGGT GGTTCCTGGCA TTTATAATTT GCTGGTTGCC CTTCCACGTT 540
GGCAGAATCA TTTACATAAA CACGGAAGAT TCGCGGATGA TGTACTTCTC TCAGTACTTT 600
AA 602

```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

Leu Val Val Val Leu Ala Phe Ile Ile Cys Trp Leu Pro Phe His Val
1      5      10      15
Gly Arg Ile Ile Tyr Ile Asn Thr Glu Asp Ser Arg Met Met Tyr Phe
20      25      30
Ser Gln Tyr Phe Asn Ile Val Ala Leu Gln Leu Phe Tyr Leu Ser Ala
35      40      45
Ser Ile Asn Pro Ile Leu Tyr Asn Leu Ile Ser Lys Lys Tyr Arg Ala
50      55      60
Ala Ala Phe Lys Leu Leu Leu Ala Arg Lys Ser Arg Pro Arg Gly Phe
65      70      75      80
His Arg Ser Arg Asp Thr Ala Gly Glu Val Ala Gly Asp Thr Gly Gly
85      90      95
Asp Thr Val Gly Tyr Thr Glu Thr Ser Ala Asn Val Lys Thr Met Gly
100      105      110
Arg Lys Trp Ser Arg Arg Gly Ser Lys Asp Ala Cys Leu Gln Ser Ala
115      120      125
Pro Pro Gly Thr Ala Gln Thr Leu Gly Pro Leu Pro Leu Leu Ala Gln
130      135      140
Leu Trp Ala Pro Leu Pro Ala Pro Phe Pro Ile Ser Ile Pro Ala Ser
145      150      155      160
Thr Arg Arg Gly Gly Gly Ser Gly Ile Tyr Asn Leu Leu Val Ala Leu
165      170      175
Pro Arg Trp Gln Asn His Leu His Lys His Gly Arg Phe Ala Asp Asp
180      185      190
Val Leu Leu Ser Val Leu
195

```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12773

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C07H 21/04; C07K 14/705; C12N 15/09, 15/63; C12Q 1/68

US CL : 536/23.5, 24.3; 435/7.2, 69.1, 320.1; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5, 24.3; 435/7.2, 69.1, 320.1; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,712,253A (LARTEY et al) 27 Jan. 1998, column 18, lines 40-56	1
X	MCKEE, K. K. et al. Cloning and characterization of Two Human G Protein-Coupled Receptor Genes (GPR38 and GPR39) Related to the Growth Hormone Secretagogue and Neurotensin Receptors. Genomics, 1997, Vol. 46, pages 426-434, see whole document.	1-6, 8
X,P	Database GenEmbl, No.AF082210, PALYHA, O. C. et al. 'Orphan G protein-Coupled Receptor from Teleost Fish Spheroides Nephelus Related to Growth Hormone Secretagogue Receptor,' Sequence listing, September 1998.	7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 22 JULY 1999	Date of mailing of the international search report 07 OCT 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer NIRMAL S. BASI Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12773

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

N-GENSEQ-34, GENEMBL, EST, SWISSPROT-36, SPTREMBL-8, APS, EMBASE, BIOSIS, MEDLINE, WPIDS, JAPIO, CAPLUS

Search terms: SEQ. ID. NO:1-7, motilin receptor, g protein coupled receptor

Form PCT/ISA/210 (extra sheet)(July 1992)*